

La bioinformàtica en l'estudi de la diversitat genètica

05/2008 - **Biologia.** La variació genètica és clau per entendre la gran diversitat d'éssers vius que poblen la Terra, així com també les diferències que existeixen entre els individus d'una mateixa espècie. Aquesta tesi estudia aquest fenomen, mitjançant el desenvolupament d'eines bioinformàtiques per a l'extracció i l'anàlisi de seqüències de DNA, la generació de bases de dades de diversitat genètica i la prova d'hipòtesis a partir de seqüències de diferents espècies i regions del genoma.



La variació genètica és la matèria prima de l'evolució: són les petites diferències entre els individus d'una mateixa població les que, amb el temps, generaran tota la diversitat d'éssers vius existent a la Terra. La descripció i explicació de les forces que controlen la variació genètica dins i entre poblacions és el tema central de la genètica de poblacions. L'obtenció d'un número explosiu de seqüències de DNA a diferents gens i espècies ha canviat radicalment les perspectives de la genètica de poblacions, transformant-la des d'una ciència empírica insuficient fins a un esforç interdisciplinari de gran abast, on els aparells de generació de noves seqüències a gran escala s'integren amb eines bioinformàtiques per a l'extracció i gestió de dades, juntament amb avançats models teòrics i estadístics per a la seva interpretació.

Aquesta tesi és un projecte de bioinformàtica i genètica de poblacions complet, l'objectiu principal del qual és l'estudi de la diversitat genètica a les poblacions. S'ha dut a terme en tres passos seqüencials: (i) el desenvolupament d'eines per a l'extracció, processat, filtrat i control de qualitat de seqüències de DNA, (ii) la generació de bases de dades de coneixement a partir de les dades obtingudes a la primera part i (iii) la prova d'hipòtesis que requereixen de dades de varies espècies i regions del genoma.

A la primera part de la tesi hem desenvolupat PDA (Pipeline Diversity Analysis, <http://pda.uab.cat>), una aplicació Web de codi obert que permet l'exploració de la variació genètica a partir dels milions de seqüències que hi ha emmagatzemades a les principals bases de dades públiques. Totes les dades extretes i analitzades a la primera part de la tesi són utilitzades a la segona part per a crear un recurs via Web de la variació genètica en el gènere *Drosophila* (DPDB, *Drosophila* Polymorphism Database, <http://dpdb.uab.cat>). Finalment, la tesi inclou dos estudis que utilitzen els mòduls d'extracció i anàlisi de dades desenvolupats a la primera part. En el primer, hem estudiat patrons de variació genètica per a inferir l'acció de la selecció natural a seqüències de DNA que no codifiquen cap gen però que en canvi estan conservades entre diferents individus d'una mateixa o de diferents espècies. Els resultats mostren que aquestes seqüències conservades no codificadores són mantingudes per la selecció purificadora a *Drosophila*. El segon estudi es centra en l'evolució dels gens Hox, una classe de gens que estan involucrats en l'especificació de les regions al llarg de l'eix anteroposterior del cos i que són per tant essencials en el desenvolupament primerenc. Sorprenentment, els resultats mostren que els gens Hox evolucionen més ràpidament que altres gens que no són essencials durant el desenvolupament o que presenten patrons d'expressió menys complexos.

Resumint, els treballs presentats a aquesta tesi tanquen un cicle complet de projecte bioinformàtic, incloent tots els passos necessaris des de l'extracció de dades fins a la generació de nou coneixement científic. És més, el resultat de cada pas és la llavor per a múltiples possibles estudis en el següent pas, i per tant aquesta tesi té moltes aplicacions per a la comunitat científica.

Figura 2. Portada de la tesi doctoral presentada per Sònia Casillas Viladerrams.

Sònia Casillas

Departament de Genètica i de Microbiologia

Universitat Autònoma de Barcelona

"Development and application of bioinformatic tools for the representation and analysis of genetic diversity". Tesi doctoral presentada per Sònia Casillas Viladerrams el 22 de febrer de 2008 a la Facultat de Biociències de la UAB i dirigida pels doctors Antonio Barbadilla i Alfredo Ruiz.